BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Design Considerations for Very-High-Speed Si-Bipolar IC's Operating up to 50 Gb/s

H.-M. Rein, *Member, IEEE*, and M. Möller

Abstract—**In this paper, design aspects of high-speed digital and analog IC's are discussed which allow the designer to exhaust the high-speed potential of advanced Si-bipolar technologies. Starting from the most promising circuit concepts and an adequate resistance level, the dimensions of the individual transistors in the IC's must be optimized very carefully using advanced transistor models. It is shown how the bond inductances can be favorably used to improve circuit performance and how the critical on-chip wiring must be taken into account. Moreover, special modeling aspects and ringing problems, caused by emitter followers, are discussed. An inexpensive mounting technique is presented which proved to be well suited up to 50 Gb/s, the highest data rate ever achieved in any IC technology. The suitability of the design aspects discussed is confirmed by measurements of digital circuits and broadband amplifiers developed for 10 and 40 Gb/s optical-fiber links.**

I. Introduction

SILICON bipolar circuits are just penetrating speed ranges which some years ago still seemed to be reserved for compound semiconductors (e.g., GaAs). For example, today there is no question that all circuits in 10 Gb/s optical-fiber links can be fabricated in available advanced Si *production* technologies (e.g., [1] and [2]). The same is true for most circuits in wireless applications with typical frequencies between 0.9 and 2.5 GHz or even 5.8 GHz. Recently, silicon MUX and DEMUX IC's were realized in a laboratory technology operating above 40 Gb/s without needing sophisticated processes [3], [4]. Si-based IC's for even higher operating speeds are just under development.

An important precondition for this speed improvement is the progress in silicon bipolar technologies. However, digital circuits with clock frequencies as high as the transistor transit frequency f_T or broadband amplifiers with cutoff frequencies up to about $f_T/2$ are only possible if adequate circuit concepts are used and if the circuits are carefully designed. Such a design does not only include the optimization of the circuit resistances (and thus of the transistor operating points and voltage swings) but also the individual design of all transistors in the circuit as well as the optimization of on-chip wiring and even of the length of the bond wires. All these aspects are discussed in this paper. The basic steps in high-speed circuit design are summarized in Fig. 1, starting with the choice of optimum circuit principles and finishing by simulating the complete IC with all parasitics.[1]

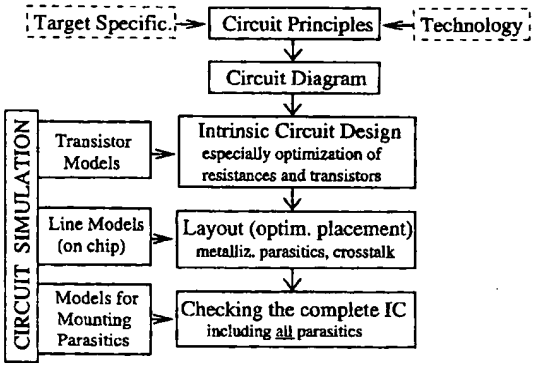


Fig. 1. Basic steps in the design of high-speed circuits. For simplification the iteration loops between the different steps are not shown.

The circuit examples discussed in this paper are digital IC's and broadband amplifiers, e.g., used in optical-fiber links. For this application, maximum speed is often a must, while the power consumption is not a limiting condition. However, most of the circuits are of interest for other applications, too. Moreover, the design aspects presented here are also applicable to low-power high-speed circuits (as shown by a single example) and most of them also to other kinds of circuits (e.g., for wireless applications).

Numerous analog and digital high-speed IC's have been designed by the authors' group (RUB) considering the aspects discussed in this paper. In Table I, some examples developed in cooperation with our industrial partners (given in parentheses) and fabricated in production technologies of different companies are summarized, together with the corresponding references. In the last column, the most recent results based on a laboratory technology are given. All these circuits achieve record values for silicon and some of them are even faster than the fastest compound semiconductor IC's. Some of the most critical IC's (with respect to achieving the demanded specifications) are shortly discussed at the end of this paper.

II. Si-Bipolar Technologies

Today, the most promising production technologies are based on self-aligning double-polysilicon processes. The

Manuscript received October 1, 1995; revised February 23, 1996. This work was supported in part by the European Union (R&D Programs RACE and ESPRIT) as well as by the German Ministry BMBF (Project Photonik II).

The authors are with Ruhr-University Bochum, D-44780 Bochum, Germany.

Publisher Item Identifier S 0018-9200(96)05707-1.

[1] In contrast to the simplified scheme in Fig. 1, design experts often use preliminary values for metallization and mounting parasitics already in the first optimization step of the circuit in order to reduce the number of iterations.

TABLE I
HIGH-SPEED IC'S FOR OPTICAL COMMUNICATIONS IN Si-BIPOLAR PRODUCTION TECHNOLOGIES—STATE OF THE ART. RECORD SPEEDS IN A LABORATORY TECHNOLOGY ARE GIVEN IN THE LAST COLUMN (SIEMENS + RUB)

Circuit	Max. Speed / Prod. Technology		Remarks	Labor. Techn.
Multiplexer	30 Gb/s (RUB + HP) / HP 25	[5]	2:1	50 Gb/s [4]
Decision Circuit	22 Gb/s (RUB + Siemens) / B6HF	[6]	CPM = 180°	25 Gb/s [12]
Demultiplexer	32 Gb/s (RUB + Siemens) / B6HF	[6]	1:2	46 Gb/s [3]
Static Frequency Divider	19 GHz (RUB + Siemens) / B6HF	[7]	2:1, 2 Outputs (0° and 90°)	30 GHz [3]
Laser Driver	14 Gb/s (RUB + Siemens) / B6HF	[8]	Swing ΔV_Q = 3.6 V (50 Ω)	20 Gb/s [22]
Preamplifier (Transimp. Type)	13 Gb/s (RUB + ANT + Siemens) /B6HF	[9]	Transimp. Z_{TI} = 615 Ω f_C = 10 GHz Noise $\bar{i}_N$ = 10.5 pA/$\sqrt{Hz}$	20 Gb/s [22]
Main Amplifier, AGC	13 Gb/s (RUB + ANT + Motorola) /Advanced MOSAIC V	[10]	S_{21} = 37 dB, f_C = 10 GHz Dynamic Range = 41 dB	—
Main Amplifier, limiting	15 Gb/s (RUB + ANT + Siemens) /B6HF	[11]	S_{21} = 52 dB (max.) Dynamic Range = 46 dB	

main advantage of double-polysilicon compared to single-polysilicon technologies is the lower base resistance. A typical example of an advanced production technology is the B6HF of Siemens [13]. It is characterized by an f_T of about 25 GHz, low base resistance, 0.8 μm lithography (resulting in 0.4 μm emitter stripe width), three metallization levels, and a rather conservative isolation technique. Most of the circuits presented in Table I are realized in this production technology or in an improved but unsophisticated laboratory version [14]. Even higher operating speeds can be achieved by use of SiGe-base technologies (e.g., [2]), as, e.g., shown by our current IC designs for a 40 Gb/s optical system [22].

For choosing the right technology for high-speed IC's, it is highly recommended not only to take a look at maximum f_T. A slightly better "figure of merit" is a simplified relation for the maximum oscillation frequency f_{max} which, in addition, contains the base resistance r_B and the total collector-base capacitance C_{CB} (consisting of junction and oxide capacitance)

$$f_{max} = \sqrt{f_T/8\pi r_B C_{CB}}. \quad (1)$$

Instead of using a single but inadequate figure of merit, the experienced designer prefers to take a look at the individual (area- and length-) *specific* electrical parameters in order to judge a technology. Especially the base resistance and emitter contact resistance, as well as the various junction and oxide capacitances, are considered (all *related* to the emitter size) and, of course, the total transit time τ_f.[2]

One of the weak points of f_{max} as a figure of merit is that the current carrying capability, given by the maximum admissible collector current density j_{CK} (see Section IV-B), is not considered. As a consequence, record values for f_{max}, achieved by weakly doped and thick epitaxial collector regions [16], are dearly paid by a low j_{CK} which is not adequate for most applications. Note that j_{CK} is a very important transistor parameter, since the high-speed *driving* capability of the transistor is strongly influenced by $j_{CK} \cdot A_E/C_{CB}$ (A_E = emitter area). In practice, j_{CK} is increased by a selective implantation in the collector (SIC) region below the emitter which additionally can increase f_T also at medium current densities.

[2]From these parameters, the designer can derive other figures of merit which are tailored to a *specific* class of circuits.

III. HIGH-SPEED CIRCUIT PRINCIPLES

A. Digital Circuits

Fig. 2 shows the circuit diagram of a master-slave D-flipflop (MS-D-FF) as an example for a frequently used circuit. It consists of the circuit core with data input stages, a clock buffer, and an output buffer for driving matched 50-Ω transmission lines. This circuit equals the decision circuit in Table I and is also the basic component of the frequency divider and the demultiplexer given there. From this diagram, most of the circuit principles recommended for high-speed IC's can be taken which shall be summarized briefly.

- *E^2CL instead of ECL.* This means that the emitter followers (EF) are used at the data inputs and the limiting differential stages, called current switches (CS), at the output. Among other advantages, this circuit concept shows essentially improved behavior in a transmission line environment compared to ECL (e.g., better line matching and steeper pulse edges).
- *Series Gating.* Besides high-speed logic connection (e.g., between data and clock signals) this concept allows differential operation.
- *Differential Operation.* This principle results in many advantages, which are summarized in Fig. 3.
- *Low Voltage Swing.* This measure increases speed (especially of the CS) and reduces power consumption. Optimum values for *internal differential* swings of the data pulses are usually about 2 × 200 mV_{p-p} = 400 mV_{p-p}.
- *Multiple Emitter Followers.* At high operating speed, the decoupling capability (impedance transformation) of emitter followers is rather limited due to the reduced effective current gain of the transistors. Thus two or even three cascaded emitter followers are often required. Moreover, by this measure V_{CE} across the CS transistors

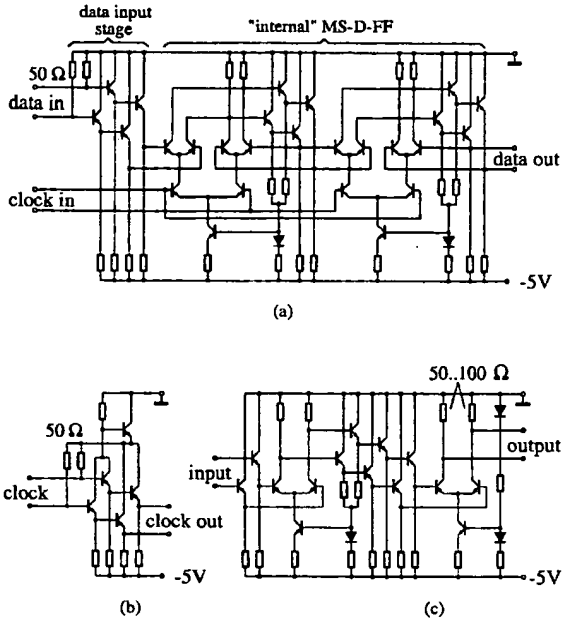


Fig. 2. Circuit diagram of a high-speed master-slave D-flipflop (MS-D-FF). (a) MS-D-FF core with data input stage. (b) Clock buffer. (c) Data output buffer.

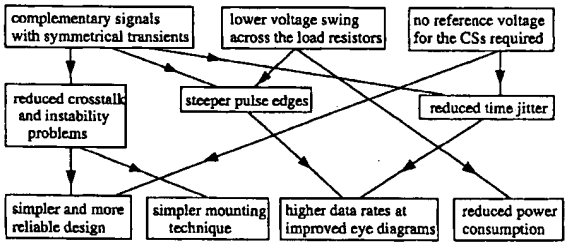


Fig. 3. Advantages of differential operation (compared to single-ended operation).

and thus the maximum collector current density j_{CK} are increased. As a consequence the transistor size can be reduced, further reducing C_{CB}.

- *On-Chip Matching* of external transmission lines, especially at the data input. Compared to off-chip matching, the return loss is essentially improved and the potential instability of emitter-follower inputs can be reliably suppressed [17], [18].

B. Broadband Amplifiers

Amplifiers in broadband communications must be operated down to very low frequencies (e.g., down to about 30 kHz in optical-fiber links). Thus, in monolithic integrated amplifiers dc coupling is necessary. Moreover, it is recommended to apply the principle of strong impedance mismatching between succeeding stages. As a typical example, a chain of alternating transadmittance stage (TAS), transimpedance stage (TIS), and

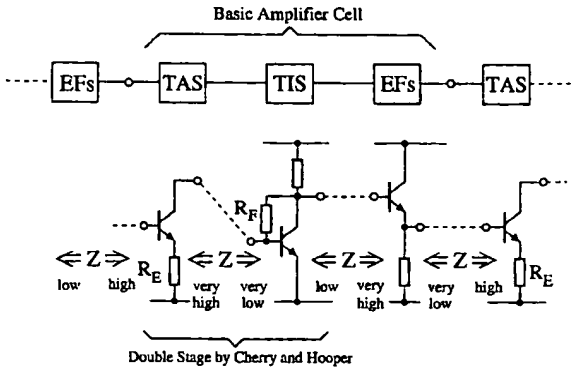


Fig. 4. Principle of strong impedance mismatching in broadband amplifiers. Z stands for the input and output impedances, respectively.

emitter followers (EF's) is shown in Fig. 4.[3] Due to this mismatching, the transfer functions of the different stages (e.g., transadmittance and transimpedance) remain approximately constant up to comparatively high frequencies, resulting in a high cutoff frequency. This is because of the drastically reduced influence of the strongly frequency-dependent input and output impedances of the different stages. As another advantage of mismatching, all nodes in the circuit are low-ohmic. This fact reduces the influence of parasitic capacitances and thus further increases the bandwidth. In Fig. 4 the original double-stage proposed by Cherry and Hooper [19], consisting of a TAS and a TIS, is extended by (one to three) EF's between the output of the TIS and the input of the TAS. These EF's are required for improving the insufficient mismatching at this interface, for level shifting, and for gain peaking near the upper frequency limit. At high frequencies, where mismatching is degraded, conjugate complex input and output impedances can be observed at the interfaces between different stages, which can further increase the bandwidth.

In the receivers of optical links (without an optical-fiber amplifier in front of the photodiode) we need a low-noise preamplifier behind the photodiode, succeeded by a high-gain main amplifier. Both circuits must be designed for a wide dynamic range. Usually, they are realized on separate chips connected by 50-Ω transmission lines [20]. In the preamplifier, the first stage of the chain of mismatched stages is generally a TIS in order to guarantee broadband mismatching between the driving (high-impedance) photodiode and the amplifier input. This stage is followed by emitter followers and an output buffer (TAS) with external 50-Ω loading [9], [21]. Using EF's at the output is not recommended (as in the case of digital circuits).

The main amplifier must have a high gain at 50-Ω input impedance. Therefore, often on-chip matching and a single EF pair are used at the input, followed by several basic amplifier cells (cf. Fig. 4) and a TAS output buffer [10], [11]. Since the main amplifier must have a constant output amplitude,

[3]Sometimes the cutoff frequency of the gain magnitude is increased by using an EF in the feedback path of the TIS [23]. However, by this measure the group delay is often degraded, even resulting in a reduction of the maximum *usable* operating speed (cf. Section IV-B).

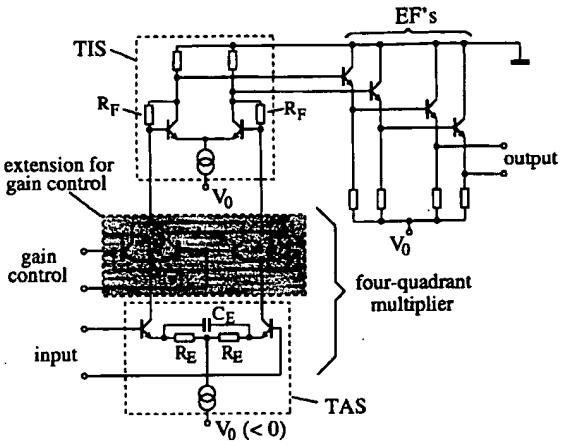


Fig. 5. Differential broadband amplifier cell. The shaded circuit part is required for AGC amplifiers only. The realization of the adjustable peaking capacitance C_E is, e.g., given in [40].

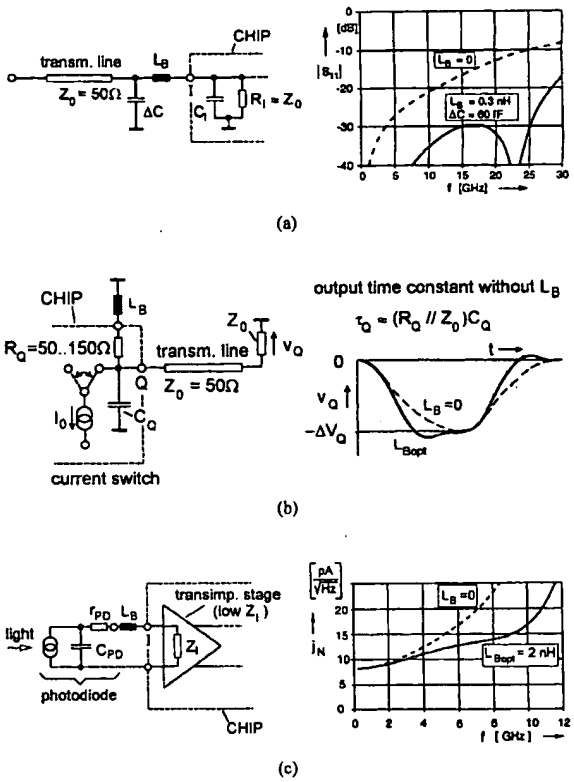


Fig. 6. Examples for favorable use of bond inductances. (a) Transmission line matching at IC input. (b) Steepening of pulse edges at IC output. (c) Improvement of signal-to-noise ratio (preamplifier).

independent of the input signal amplitude, two solutions are possible: limiting amplifier or automatic-gain-control (AGC) amplifier. If possible, differential operation should be preferred (cf. Section III-A, Fig. 3). A differential version of the basic amplifier cell in Fig. 4 is shown in Fig. 5. Without the shaded circuit part, it has a constant gain at small input amplitudes, but it operates also excellently in the limiting range and is, therefore, well suited for limiting amplifiers (but often with $R_E = 0$) [11]. For AGC amplifiers the gain of the amplifier cell must be controllable. To our experience, the best results for such amplifiers with respect to flat gain response and high cutoff frequency *over a wide dynamic range* are obtained if the TAS is extended (by the shaded circuit block in Fig. 5) to a four-quadrant multiplier. In this example, the gain is controlled by a dc voltage at the upper input.

Broadband amplifiers operating near the limit of a given technology are more sensitive to fabrication spread and design uncertainties (e.g., errors in modeling of the transistors as well as of the metallization and mounting parasitics) compared to digital circuits. Moreover, they are more sensitive to changes in the circuit environment (e.g., to the driving source impedance). Therefore, it is highly recommended to look for measures which enable us to correct the frequency response after fabrication or, if possible, even after chip mounting. This especially holds if the circuits are designed in an early stage of technology and system development, as for most of our circuits.[4] One example is to adjust the gain peaking capacitance C_E in Fig. 5. This is possible by cutting links of a multicapacitor network in the upper metallization level (by a laser or ultrasonic cutter). Another very simple way, but with a limited capacitance range, is to realize C_E by junction capacitances which can be varied by adjusting the dc bias voltage via external potentiometers [40]. We often combine both measures. In addition to C_E we sometimes use a narrow-band peaking network at the emitter node of the TAS [9]. It consists of a junction capacitance in series to the (inductive) output impedance of cascaded emitter follower stages. This circuit is well suited to adjust the magnitude of the gain and the group delay *near the upper frequency limit* of the amplifier. Moreover, for fine-adjustment of the frequency response, it is sometimes useful to vary (via an external potentiometer) the current through that EF which drives the TAS.

[4] In a later stage, if the data of system and technology are well established, this option of frequency-response correction may no longer be necessary.

C. Utilization of Bond Inductances for Improving Circuit Performance

A certain disadvantage of present Si technologies compared to compound semiconductor technologies is that adequate on-chip inductors at frequencies above 5 GHz are usually not available. Thus we have to think about improvement of circuit performance by applying bond wires as inductors [1]. Some examples for the favorable use of bond inductances in high-speed digital and broadband amplifier IC's are given in Fig. 6.

Good matching of the driving transmission lines, especially at the data inputs, is a must in many high-speed IC's. Otherwise, double-reflections can increase the time jitter considerably [17], [18]. Fig. 6(a) demonstrates how this matching is improved by the bond inductance L_B. In this example the circuit input, which is roughly approximated by an RC network, is driven by a microstrip line with the characteristic impedance $Z_0 = 50\ \Omega$. $R_I \approx Z_0$ is the on-chip matching

resistance (cf. Section III-A) and C_I represents the total input capacitance of the IC, consisting of bond pad capacitance and averaged input capacitance of the first (loaded) EF. Mostly we use the minimum possible bond inductance $L_{B\,\min}$. Even then, typically $\sqrt{L_{B\,\min}/C_I} > Z_0$ holds. Then a capacitance ΔC is added to the end of the line[5] so the condition

$$\sqrt{\frac{L_{B\,\min}}{C_I + \Delta C}} \approx Z_0 \tag{2}$$

is approximately met. The LC network built by ΔC, L_B, and C_I acts roughly as a lengthening of the line up to the on-chip matching resistor. Of course there is an upper frequency limit of this matching method due to the cutoff frequency of the (lumped) LC network and, moreover, due to the fact that the input of the circuit can no longer be modeled by a constant capacitance at high frequencies. It should be noted that the situation is simplistically presented in Fig. 6(a) (cf. [18]). For example, at single-ended operation, the ground bond inductance [not shown in Fig. 6(a)] influences the input impedance of the circuit and must also be taken into account [10]. As a practical example, an input return loss of $S_{11} < -20$ dB up to 11 GHz was measured for a limiting amplifier fabricated in a production technology [11].

The right-hand side of Fig. 6(a) shows a simulated example based on an advanced laboratory technology. Here, the return loss at the data input of a high-speed digital IC [cf. Fig. 2(a)] is plotted versus frequency. In this case, $C_I \approx 60$ fF and $L_B = 0.3$ nH. From (2) we get $\Delta C \approx 60$ fF. The result shows that there is a good chance to obtain a fairly good line matching up to 30 GHz, which is sufficient for data rates up to about 40 Gb/s. As shown by the dashed line, S_{11} is much worse without the $L_B - \Delta C$ network.

Fig. 6(b) demonstrates steepening of pulse edges by a peaking bond inductance. Here, the output stage of a digital IC is approximated by a switched current source (I_0) with the output capacitance C_Q and the on-chip output resistance R_Q. In this case, R_Q is connected to ground via a bond wire with inductance L_B. As a consequence, fall and rise times are reduced compared to the value determined by the output time constant τ_Q, which is given by the effective load resistance ($R_Q//Z_0$) and the output capacitance C_Q. This is shown by comparison of the solid pulse shape for optimum $L_B(= L_{B\text{opt}})$ with the dashed shape for $L_B = 0$. This measure is of special importance for driver circuits with high output voltage swing ΔV_Q, since there τ_Q is roughly proportional to ΔV_Q. For example, it was required to meet the demanded specifications of the laser/modulator driver in Table I (cf. Section VII-A and [8]).

In Fig. 6(c) it is shown how the signal-to-noise ratio (determined by the reciprocal value of the equivalent input noise current density j_N) can be increased in the case of a transimpedance preamplifier by use of an (overlength) bond wire [9], [21], [24]. For this, the input signal amplitude is peaked near the cutoff frequency f_C of the amplifier by the parallel resonance circuit consisting of L_B and the capacitance C_{PD} of the photodiode (PD). Note that damping of this resonance circuit is limited due to the comparatively low values of the input impedance Z_I of the amplifier and of the series resistance r_{PD} of the PD. A very rough approximation for the optimum value of L_B is given by

$$L_{B\text{opt}} \approx [(2\pi f_C)^2 C_{PD}]^{-1}. \tag{3}$$

Moreover, it is shown in Fig. 6(c) how L_B reduces j_N at high frequencies in a typical preamplifier for a 10 Gb/s optical system [21]. Note that here j_N is related to the *internal* node of the photodiode.

IV. Intrinsic Circuit Design

To limit the length of this paper, most of the examples given in this section are restricted to digital circuits. However, some of the basic ideas are also valid for analog circuits.

A. Resistance Level and Power Consumption

Usually the power consumption P of very high speed IC's with comparatively small-scale integration (SSI) primarily depends on whether or not the (high-speed) outputs of the circuits are heavily loaded and only in the second place by the high operating speed. The situation is qualitatively expressed by Fig. 7, which shows the basic relation between operating speed and power consumption for strong and weak loading, respectively. Strong loading, e.g., by 50-Ω transmission lines,[6] increases the power consumption drastically not only by the need of powerful output buffers but also by the increased power consumption of the circuit core and input stages (cf. Fig. 2). The main reason for this fact is the severely reduced decoupling capability of the EF's at high speed.[7] In order to improve decoupling between external load and internal circuit nodes, power-consuming cascaded EF's and sometimes (at the speed limit) even two current-switch stages are required for the output buffer [cf. Fig. 2(c)]. Even then the loading of the internal circuit nodes is still comparatively high. Therefore, these nodes must be low-ohmic resulting in an increased power consumption of the circuit core and the input stages, too. Again, the power consumption of these circuits is further increased due to the reduced decoupling capability of the EF's [cf. Fig. 2(a) and (b)].

In contrast, if the output of a high-speed circuit is only weakly loaded (e.g., by similar circuits on the chip rather than by 50-Ω buffers) the power consumption can be drastically reduced (Fig. 7). Of course, if the circuit has to be operated near the speed limit of the technology, the last, e.g., 20% in speed has to be paid by an overproportional increase of power consumption. In order to confirm this statement experimentally, first a 2:1 static frequency divider was developed on the base of a MS-D-FF and fabricated with a production technology [7]. This circuit was designed with the aim to achieve maximum speed. Despite the fact that both outputs of the internal MS-D-FF (latches) are heavily loaded by 50-Ω output buffers a maximum input frequency of 19 GHz was achieved (cf. Table I), however, at cost of a high

[5] ΔC can be realized in a simple manner by broadening the microstrip line at its end. It also includes the stray capacitance at the end of the line.

[6] This is the typical situation for high-speed SSI IC's in optical-fiber links. For these IC's power consumption is normally no limiting condition.

[7] This is because of the drastic reduction of current gain ($|\beta| \approx f_T/f$) at such high frequencies (e.g., up to $f = f_T/2$ or even above).

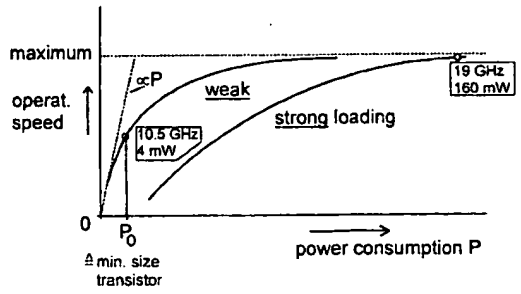


Fig. 7. Basic relation between operating speed and power consumption for different loading conditions. The power consumption of the output buffers in the case of strong loading, which would even increase the difference between both curves, is not considered. (For numerical examples and the definition of P_0 cf. text.)

power consumption of 160 mW (power consumption of the buffers not included). Then, in order to cancel the influence of peripheral loading, a 16:1 divider was developed, whose first (speed limiting) stage is only weakly loaded by the second one [7]. Now it is possible to reduce the power consumption to only 2.5% (4 mW) of the first example, still achieving a maximum frequency of 10.5 GHz, i.e., of 55%. For this, we started from the high-speed divider keeping the voltage swings constant but applying the following measures (cf. [25]).

- First, the emitter lengths $l_{E\nu}$ (ν stands for the different transistors) were reduced according $l_{E\nu} \propto P$, down to a power consumption at which the minimum emitter length given by the technology ($l_{E\,\min}$) was achieved for the upper current switch transistors [cf. Fig. 2(a)]. Note that at not too small $l_{E\nu}$ the transistor junction (+ oxide) capacitances C_ν are roughly proportional to $l_{E\nu}$ and thus to P, thus keeping the corresponding time constants $R_\nu C_\nu$ nearly constant (circuit resistances $R_\nu \propto 1/P$).[8] This also holds for the ratio $r_{B\nu}/R_\nu$.
- Next, some of the transistor configurations were changed from two base contacts to one, further reducing the collector and substrate capacitance (but, of course, increasing r_B).
- Then, the circuit diagram was simplified, especially the number of (power consuming) emitter followers was reduced. This allows to reduce the supply voltage (here from 5 to 3.5 V).

The final power consumption of 4 mW is marked in Fig. 7 by P_0. Only for $P < P_0$ where the transistor size can no longer be adjusted to the power consumption, the curve is going to approximate the well-known proportionality between speed and power (dashed line in Fig. 7). This onset of strong speed reduction with decreasing power can be shifted to lower values of P if the implantation dose of the SIC (Section II) is reduced, thus reducing C_{CB} [25].

Based on the loading considerations discussed in this section and starting with a typical (low) internal voltage swing (cf. Section III-A), we get a feeling for the resistance level and transistor operating points. Finding the optimum collector currents, especially for the different emitter followers, requires some experience and careful simulations.

In the case of analog circuits, e.g., broadband amplifiers, additional aspects for resistance optimization have to be considered. For example, in the TIS of a preamplifier the feedback resistance R_F is a compromise between low noise (→ high R_F) and high bandwidth (→ reduced R_F) [21]. There is also an optimum for the collector current of the TIS transistor with respect to minimum broadband noise and maximum cutoff frequency, respectively. In the main amplifier cell of Fig. 5 the optimum choice of the emitter series resistance R_E is a compromise between wide dynamic range and high bandwidth on one side (→ high R_E) and low noise (first stage) as well as high gain on the other side (→ low R_E). The optimum choice of R_F in this cell is mainly a tradeoff between high gain (→ high R_F) and high bandwidth (→ lower R_F) [10]. But it should be mentioned that at given total gain of a multistage amplifier, increasing the bandwidth of the single cells by reducing R_F requires more cells and thus may again reduce the bandwidth of the complete amplifier.

B. Optimization of the Transistors in High-Speed Circuits

There is a great degree of freedom for the designer of monolithic integrated circuits in adjusting the transistor dimensions (at given operating points) with respect to optimum circuit performance. Regrettably, this opportunity is often not exhausted by the designer.

Before starting the transistor optimization, the minimum admissible emitter area must be calculated in dependence on the operating point (I_C, V_{CE}). It is determined by the critical collector current density j_{CK} (cf. Section II) above which a strong increase of the transit time τ_f (decrease of f_T) can be observed due to high-current effects ("Kirk-effect" [26]). Therefore, this region is normally not allowed for high-speed transistor design (for exceptions see e.g., [8]). The value of j_{CK} can be taken from measurements, from analytical relations (e.g., [27], [28]),[9] or from a mixture of both. It increases with increasing internal collector-emitter voltage V_{CEi},[10] as well as increasing doping concentration N_C and decreasing thickness w_C of the collector region below the emitter, and it decreases with increasing junction temperature ϑ_j (e.g., [27], [28]). An experimental example for the dependence of τ_f on the averaged collector current density $j_C = I_C/A_E$ (A_E = emitter area) is given in Fig. 8 with V_{CEi} as a parameter. At small emitter width b_E the increase of τ_f with j_C is mitigated due to current spreading (e.g., [28] and [29]). As a consequence, the designer can use somewhat higher values for j_{CK} compared to the wider emitter stripes. A corresponding correction factor is proposed in [29].

From the worst-case (w.c.) values of V_{CEi} and ϑ_j and considering the fabrication spread (especially of N_C and w_C) we get the w.c. value for j_{CK}. From this, with the w.c. value

[8] However, it should be mentioned that the on-chip wiring and bondpad capacitances cannot be reduced correspondingly.

[9] Per definition, j_{CK} in [27] is slightly higher compared to the value in the present paper which is given by the *onset* of the high-current region.

[10] Compared to the terminal voltage V_{CE}, the internal voltage V_{CEi} is reduced by the voltage drop across both the external collector resistance and emitter resistance.

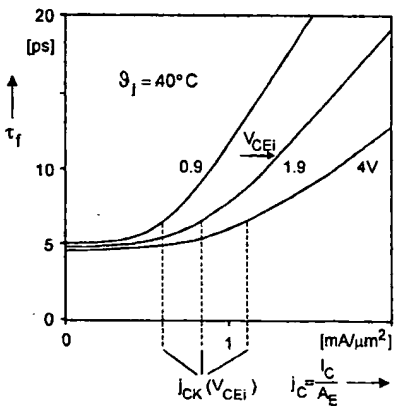


Fig. 8. Transit time versus collector current density with the internal collector-emitter voltage V_{CEi} as a parameter.

of I_C, the minimum admissible emitter area can be calculated

$$A_{E\min} = I_C/j_{CK}, \quad \text{with } j_{CK} = f(V_{CEi}, \vartheta_j, N_C, w_C)_{\text{w.c.}} \tag{4}$$

For a given emitter stripe width b_E the condition for the *total* emitter length[11] is $l_E \geq A_{E\min}/b_E$.

In most cases the optimum b_E equals the minimum value b_{E0} given by the technology because of the demand on a low base resistance (exceptions see below). Thus, starting from $l_E = A_{E\min}/b_{E0}$ the emitter length can be increased in order to find the optimal transistor size for circuit performance. This optimum is a compromise between the contradicting demands on low base and emitter resistance r_B and r_E (i.e., large l_E) on one side and low junction (+ oxide) capacitances C_{EB}, C_{CB}, and C_{CS} (i.e., small l_E) on the other side. Usually, in high-speed circuits the dominating transistor parameters, besides τ_f, are r_B and C_{CB}.[12] Their dependence on l_E (at $b_E = b_{E0} = \text{const.}$) can be roughly expressed by

$$r_B \propto 1/l_E \quad \text{and} \quad C_{CB} \propto l_E \tag{5}$$

provided l_E is not too small. As a rough rule of thumb, in digital circuits with a configuration similar to that in Fig. 2, the transistors in upper current switches (CS) are usually optimal for minimum emitter lengths, i.e., $l_E = A_{E\min}/b_{E0}$ (provided that the length-specific base resistance is not too high). This is mainly due to the strong influence of C_{CB} caused by the Miller effect and the dotting of *two* collectors which doubles the capacitive loading by C_{CB} and C_{CS}. Minimum length is often also optimal for the first one of cascaded EF's since this transistor is already sufficiently long due to its comparatively low j_{CK} (low V_{CEi}) and because its output is rather weakly loaded. Moreover, a low C_{CB} is desired since it either increases the loading of the upper internal CS or degrades transmission line matching at the circuit input (cf. Section III-C). In contrast, the second EF in double-EF stages is usually longer than its minimum value because it is (strongly) loaded by a CS and needs, therefore, a low r_B (low output impedance) rather than a low C_{CB} (no Miller effect). However, it should be noted that these high-speed design considerations for EF's may fail if ringing on top of the pulses occurs which, e.g., increases the time jitter. This potential problem of EF stages often requires a tradeoff between speed and stability in the design of EF transistors (cf. Section IV-C).

In contrast to the design $b_E = b_{E0}$ discussed up to now, the choice $b_E > b_{E0}$ at $A_E = A_{E\min}$ may be advantageous for transistors whose base resistance is of minor importance while C_{CB} and C_{CS} have a strong influence on circuit performance. Here, a reduction of l_E (i.e., $l_E < A_{E\min}/b_{E0}$) and thus of C_{CB} and C_{CS} may be advantageous. However, it should be mentioned that if b_E exceeds b_{E0}, r_B increases stronger with reduced $l_E (= A_{E\min}/b_E)$ than given by (5), as for the *internal* base resistance now $r_{Bi} \propto b_E/l_E = A_{E\min}/l_E^2 = b_E^2/A_{E\min}$ holds. A typical example for the design $b_E > b_{E0}$ are transistors in output buffers with comparatively high voltage swing ΔV_Q, where the output capacitance (mainly determined by C_{CB} and C_{CS}) is of dominating influence. For example, for the laser/modulator driver in Table I, $b_E = 2b_{E0}$ was chosen [8]. There, due to the high ΔV_Q, the collector current I_C and thus $A_{E\min}$ are high, resulting in a still sufficiently low r_{Bi}. Another example where the design $b_E > b_{E0}$ may be advantageous are transistors in current sources. Again, low values of C_{CB} and C_{CS} are of main interest, while r_B is normally of smaller influence.

From these considerations, an optimization procedure for the transistors in a digital circuit can be derived (e.g., with the aim of a clear eye diagram at maximum possible data rate). The transistor dimensioning starts from a given circuit diagram (cf. Section III-A) and, additionally, it is assumed that a first set of the resistances in the circuit has already been chosen (cf. Section IV-A) so that the operating points of the transistors are given. Then the following (simplified) procedure is recommended.

- Calculate the w.c. values of $j_{CK\nu}$ (minimum) and $I_{C\nu}$ (maximum) for each transistor ("ν" stands for the different transistors).
- Calculate the minimum emitter areas $A_{E\nu,\min} = I_{C\nu}/j_{CK\nu}$.[13]
- Check whether I_C is below the maximum admissible current I_M for the metal stripe of a single emitter finger. Otherwise, increase the number of emitter fingers ($n_E \geq I_C/I_M$).
- Calculate the parameter sets of all transistors for $A_{E\nu} = A_{E\nu,\min}$ and minimum emitter stripe width b_{E0} (i.e., $l_{E\nu} = A_{E\nu,\min}/b_{E0}$). Check the eye diagrams and determine the maximum data rate. This starting dimensioning normally proves to be already the optimum for transistors in the upper CS's and often also in the lower CS's and the first EF stages.
- Lengthen the emitter stripes of the switching transistors (i.e., $l_{E\nu} > A_{E\nu,\min}/b_{E0}$) and check whether the data rate can be increased and the eye diagrams can

[11] In general, l_E is the number of emitter fingers times the single emitter stripe length.

[12] However, in future IC's the influence of r_E may strongly increase due to the ongoing reduction of b_E (e.g., [30]) combined with an increase of j_{CK}.

[13] For simplification, it is not considered here that, due to emitter-size tolerances, the *nominal* $A_{E\nu,\min}$ is slightly larger in order to meet (4) *under w.c. conditions*.

be improved, respectively. An improvement is expected especially by the second (third) EF.

- Widen the emitter stripes ($b_{E\nu} > b_{E0}$) of those transistors which are still at $A_{E\nu,\min}$ and adjust $l_{E\nu}(= A_{E\nu,\min}/b_{E\nu})$. There may be some improvements by the transistors in output buffer stages with comparatively large voltage swing, in current sources, and sometimes in the lower CS's driven by the clock.

If the demanded data rate is not achieved, an iteration loop in the design procedure of Fig. 1 is required, possibly starting from a modification of the circuit diagram and its resistances (e.g., increase of power consumption). However, if the data rate is far above the target value, the power consumption and thus the transistor sizes can be reduced.

If the EF's in the circuit are strongly capacitively loaded at comparatively low static I_C, the peak collector current $\hat{i}_C$ can exceed the static value considerably. Thus, if $A_{E\min}$ from (4) is calculated by use of I_C, the transistor enters the high-current region for a short time which might increase the time jitter at maximum data rate [31]. On the other hand, if $\hat{i}_C$ is used for calculating $A_{E\min}$ the emitter may become unfavorably long. In this case, a reliable emitter design requires transistor models which are sufficiently accurate in the high-current region. However, such models are not always available [32]. In practice, the problem is mitigated by the facts that (as already mentioned) the optimum A_E of the second (third) EF is usually larger than $A_{E\min}$ and that the collector current peak of the EF's is normally very short.

Up to now the transistor design was based on the condition $j_C \leq j_{CK}$. As mentioned before, the junction capacitances (C_{CB}, C_{CS}) in buffer stages with comparatively high output voltage swing dominate the operating speed. They can be reduced by choosing $A_E < A_{E\min}$. In this case, a certain increase of the transit time τ_F by penetrating the high-current region may be tolerable [8].

For analog circuits, some additional aspects in transistor design have to be considered. For example, in the first stage of low-noise amplifiers, the choice $l_E > A_{E\min}/b_{E0}$ may be advantageous in order to reduce r_B and r_E and thus the noise figure. Moreover, in linear broadband amplifiers for high data rates, not only the magnitude of the gain but also the group delay must be approximately constant up to the required cutoff frequency f_C. Otherwise the time jitter grows unacceptably high. A rough condition is given by [21]

$$f_C/\text{GHz} \geq 0.75 \times (\text{data rate})/\text{Gb/s}. \tag{6}$$

These conditions must be met by a careful optimization of the dimensions and thus the parameters of the transistors (e.g., [21] and [10]).

This discussion on transistor optimization shows that the model parameters are required for many quite different-sized transistors. If all these parameters have to be determined experimentally by parameter extraction methods the measuring effort is enormous. Moreover, after even a small change in technology, all transistors in the library have to be fabricated and measured again in order to get the new parameter set. Therefore, it is advantageous to use a "semiphysical" transistor model. It utilizes the (comparatively simple) relations between the model parameters (base resistance, junction capacitances, etc.) on one hand and the transistor dimensions as well as *specific* electrical and technological parameters on the other hand [33]. Such specific parameters are, e.g., the sheet resistances of the internal base, of the polysilicon layers, and of the buried layer (in $\Omega/\square$), the specific contact resistances (in $\Omega\mu m^2$ and $\Omega\mu m$, respectively), the length-specific and area-specific values for the periphery and the bottom of the junction capacitances (in fF/μm and fF/μm^2, respectively), thickness and doping concentration of the epitaxial collector, etc. All these specific parameters as well as their temperature dependence can be experimentally determined by simple large-size test structures, so that the measuring effort and error are only small. From these data, the model parameters for arbitrary transistor configurations can now be calculated without needing a voluminous transistor library. Moreover, this is the only practical way to predict the influence of process tolerances on circuit behavior by simulation. A corresponding computer program called TRADICA was developed by our group [33]. It is intensively used for circuit design, also by the industry.

C. Some Specific Design Problems

The problems discussed in this section are restricted to the *intrinsic* circuit. Further problems concerning the layout and the mounting parasitics are discussed in Section V and VI, respectively.

1) Transistor Modeling Aspects: The closer a circuit operates to the speed limit of the technology used, the stronger are the demands on the accuracy of the transistor models. Accurate models which overcome the problems of the frequently used Gummel/Poon-model in SPICE are available for Si-bipolar transistors.[14] Such a model, named HICUM (from HIgh-CUrrent Model) has been described in [27]. It is semiphysical in the sense discussed in Section IV-B and can be used up to high current densities. Its extension for advanced transistors and operating frequencies up to f_T has recently been published [35], [31], [29]. Some improvements of this "Advanced HICUM" compared to the original model shall briefly be discussed now.

In very high speed circuits, nonquasi-static effects have to be considered more accurately, since the operating frequency is no longer low compared to the *intrinsic* transit frequency of the transistor, $f_{T0} = (2\pi\tau_f)^{-1}$. This was confirmed by numerous simulations. For example, these effects degrade the frequency response and stability of amplifiers and support potential ringing of cascaded EF's even in digital circuits. This nonquasi-static behavior of the internal transistor is taken into account in the Advanced HICUM by additional delay times (τ_1, τ_2) for the emitter diffusion capacitance C_{de} and the intrinsic transconductance g_m (cf. [44])

$$C_{de} = C_{de0}\exp(-j\omega\tau_2), \tag{7a}$$

$$g_m = g_{m0}\exp[-j\omega(\tau_1 + \tau_2)]. \tag{7b}$$

[14] A critical discussion of the different models is given in [32] and [34].

As a consequence, the usual frequency dependency of the current gain β is modified by multiplication with $\exp(-j\omega\tau_1)$. Since τ_1 and τ_2 cannot be measured accurately, we determine these parameters by device simulations based on the transistor doping profile.[15]

As another improvement of the Advanced HICUM, the increasing influence of the peripheral active transistor with decreasing emitter width b_E is considered more accurately, e.g., by a reduced effective internal base resistance r_{Bi}. Moreover, the dynamic emitter current crowding is accurately modeled by shunting a capacitance to r_{Bi}.

Nowadays, there are still modeling aspects which have to be considered more carefully. As an example, the influence of a high-ohmic p-substrate is not considered with sufficient accuracy in transistor models. For example, at high frequencies, the transistor isolation box can no longer be modeled by the usual series connection of the junction capacitance C_{CS} and an ohmic series resistance (with the latter usually not known to the designer). One reason for this is the dielectric behavior of the substrate which can be considered by a capacitance shunted to the substrate resistance. Moreover, the spreading resistance of the channel stopper region must be considered carefully. An improved four-element equivalent circuit for the isolation box has been presented in [36] and [45] and confirmed by measurements up to 20 GHz. The model parameters are calculated from the layout and technological data by three-dimensional numerical simulation.

The influence of inaccurate substrate modeling on transistor behavior is usually more pronounced in analog circuits compared to digital circuits. This also holds for the parasitic coupling between the different transistor isolation boxes and other circuit components, which may, e.g., degrade the frequency response and stability of high-gain amplifiers as well as the performance of mixed-mode IC's. The development of adequate methods for the consideration of the crosstalk problems in very-high-speed circuit designs are still in a preliminary stage (for references see, e.g., [45]).

Another imperfection of most transistor models implemented in circuit simulators is that breakdown phenomena are not or not sufficiently considered. This is a severe problem especially in the design of high-speed IC's as discussed in the following section.

2) The Breakdown Problem: An improvement of silicon bipolar technologies with respect to higher operating speed must be paid by a reduction of the breakdown voltages V_{CE0} (open base) and V_{CB0} (open emitter). The fastest production technologies have, e.g., $V_{CE0} \approx 3.5$ to 5 V and $V_{CB0} \approx 12$ to 18 V. In advanced laboratory technologies these voltages can be considerably lower (e.g., $V_{CE0} \approx 2.5$ V). Thus, the designer of very-high-speed IC's must learn to handle such low breakdown voltages. In several cases, high collector-emitter voltages V_{CE} can, e.g., be simply avoided by connecting one or more emitter-base diodes in series to the collector. Problems arise if high output voltage swings are required (e.g., [8]). Fortunately, the base of a fast switching transistor in a circuit is typically driven by a low-ohmic source (rather than by a current source) and, additionally, often an emitter series impedance exists. As a consequence, the relevant breakdown voltage is higher than V_{CE0}. However, it should be mentioned that for $V_{CE} > V_{CE0}$ the base current I_B changes its sign. The corresponding voltage drop across the internal base resistance r_{Bi} results in a pinch-in of the collector current and can lead to electrical second breakdown (e.g., [37], [38]). This effect, which reduces the maximum admissible V_{CE} compared to $r_{Bi} = 0$, is the reason why for transistors in common-base configuration and in EF's the practical breakdown voltage can be essentially lower than V_{CB0}. In the case of transistors with multiple emitter stripes, it may be advantageous to use emitter balance resistors in order to obtain a uniform current distribution.

In conclusion, the usual condition $V_{CE} < V_{CE0}$ is often not justified and can unnecessarily restrict the application range of a given technology. However, for $V_{CE} > V_{CE0}$ accurate consideration of breakdown phenomena in circuit simulators, as, e.g., in [39], is highly recommended.

3) Ringing Caused by Emitter Followers: Ringing is a special problem of EF's and has, therefore, to be considered carefully during circuit design. At the circuit input, the negative real part of the EF's input impedance can, together with the input capacitance, the bond inductance, and the driving source impedance, even cause undamped oscillations. However, by use of the on-chip transmission-line matching, discussed in Section III, this problem can be reliably avoided.

More critical is the potential "on-chip" ringing or (even undamped) oscillation. It can be caused by the low inductive output impedance of the driving EF or TIS (possibly supported by an inductive behavior of the connection line, cf. Section V) and the capacitive input impedance (combined with a negative real part) of the loading EF, current switch (CS), or TAS. Mostly, the problem increases with the number of cascaded EF's and the length of the interconnection lines (cf. Section V). However, by a careful individual design of the transistors and their operating points, as well as of the on-chip wiring, these problems can be overcome, but often at cost of the pulse steepness and cutoff frequency, respectively. This also hoilds if a resistor is used in series to the base of the CS transistors.

In digital circuits with differential operation, an adequate compromise between speed and stability is sometimes possible if that EF pair which directly drives the CS has only a comparatively small quiescent current.[16] As a consequence, during switching, the off-going EF transistor is switched off ($i_C \approx 0$) for a short time, thus damping the ringing on the top of the pulse. The speed reduction by this measure is often tolerable since the peak current of the ongoing EF transistor is mainly determined by the input capacitances of the succeeding stage during the transient, rather than by its quiescent current.

As an experimental example, Fig. 9 shows a section of the (single-ended) output pulse sequence of the laser/modulator

[15] It should be mentioned that the influence of nonquasi-static effects is typically increased (compared to the values given in [31] and [35]) in advanced high-speed technologies due to the increased (relative) contribution of the travelling time through the collector space charge region and, possibly, by a drift field in the base generated by a Ge gradient.

[16] Sometimes this measure may also be advantageous in (differential) limiting amplifiers.

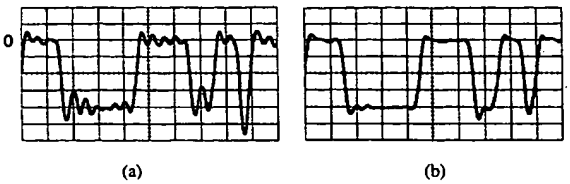


Fig. 9. (a) Ringing of the output voltage of a driver circuit caused by EF's. (b) Avoidance of ringing by the measures discussed in the text. Data rate 10 Gb/s (200 ps/div.).

driver in Table I [8]. This circuit consists of two cascaded current switches, each driven by three EF stages (cf. Section VII-A). In Fig. 9(a), a severe ringing is observed which drastically degrades the eye diagram. This ringing, which has already been detected by simulation during circuit design, is caused by the first current switch and the driving three EF's. By reducing the quiescent current of the third EF, as discussed before, the ringing is sufficiently damped without essential degradation of the pulse steepness as shown in Fig. 9(b). A similar measure was, e.g., applied in the two-stage output buffer of the decision circuits in Table I [cf. Fig. 2(c)] [6]. As an additional advantage of this method, the quiescent EF current can still be readjusted after chip mounting by simply using an external resistor.

V. Layout Aspects

For differentially operating digital circuits and broadband amplifiers, a highly symmetrical layout should be aimed at. By this measure, the advantage of differential operation with respect to drastically reduced electrical and thermal interactions (via the p-substrate and the bond wires) as well as small time jitter (cf. Fig. 3) can be fully exhausted. In addition, a large ground contact (e.g., realized as a wide ground ring around the chip) with a sufficient number of bond pads is recommended. This measure reduces the critical parasitic ground inductance caused by on-chip wiring and bond wires, which may result in common mode noise.

In single-ended amplifiers, the situation can be quite different. For example, interactions via a common ground bond wire must often be reduced by using a separate bond wire for the ground of the first amplifier cell. While *on-chip* decoupling of the supply voltage can simplify the *off-chip* decoupling in differential amplifiers (e.g., [10]) it is often not recommended for single-ended amplifiers. This is due to the influence of a parasitic resonance circuits built by the on-chip decoupling capacitance and parasitic inductances (caused by both bond wires and on-chip wiring). But even if no extra on-chip decoupling capacitance is applied, a similar effect can be caused by the parasitic capacitance between ground metallization and substrate (possibly increased by the collector-substrate capacitance of the EF's), especially if a low-ohmic (p^+) substrate is used. Thus, minimization of both the parasitic inductances and capacitances (tradeoff) is necessary in order to shift the corresponding resonance frequency far above the operating frequency of the amplifier. As another problem, mainly at single-ended operation, potential noise coupling via the substrate must be considered. In principle, this coupling can be reduced by shielding. However, especially in the case of sensitive circuit nodes, we must carefully check whether this measure is really helpful or, on the contrary, even degrades circuit performance (e.g., due to a noisy shield potential).[17]

Another very important layout aspect is a careful design of signal lines on the chip. These lines can influence the circuit performance severely, a fact which is often underestimated. Since there is, for several reasons, a lower limit for the chip size (even for small-scale IC's) the signal path(s) between the input(s) and output(s) can usually not be chosen as short as desired. Maximum line lengths of several 100 μm are often unavoidable. For these reasons, accurate line modeling is a must. Therefore, we prefer to realize longer lines as microstrip lines with the signal stripe in one of the upper metallization levels and the ground plane in the lowest level. Such a line is fairly well defined (depending on the "quality" of the ground plane) and can be modeled accurately also by lumped elements. In contrast, modeling of lines running over the substrate is very complicated since the model parameters do not only depend on the inhomogeneous nature of the underlying substrate but also on neighboring regions (channel stopper, transistor boxes, substrate-contact and other metallized areas, etc.) and, moreover, on the line length.

With the same argument of prime priority for accurate circuit simulation, we often use the buried layer as a grounded plane underneath signal pads (but, see footnote 17). As another advantage, the only small resistance (caused by the buried layer) in series to the pad capacitance may simplify the matching of the driving transmission line, discussed in Section III-C [Fig. 6(a)].

Optimum partitioning of a high-speed circuit may depend on thermal and crosstalk considerations. Another (often the dominating) partitioning aspect is to divide the total signal-line length between input and output pads into optimum sections. For this, we must know which lines can have a strong (normally negative) influence on circuit behavior ("critical" lines) and must, therefore, be considered carefully, and which lines have only a small influence ("uncritical" lines). As a rough rule of thumb, normally the critical lines should be shortened at cost of the uncritical ones. As an example, the different line segments in the signal path of a digital circuit and of a broadband amplifier with respect to their influence on circuit behavior are characterized in Fig. 10. Uncritical are lines driven by a CS[18] or a TAS due to their high output impedance. In contrast, critical are lines driven by EF's or a TIS due to their low (inductive) output impedance and to the capacitive input impedance (usually combined with a negative real part) of a loading EF, CS, or TAS. The influence of these lines is most pronounced in low-ohmic (e.g., driver) circuits where the EF's are designed for a very low output impedance (low r_B and high I_C). In this case, even small

[17] For example, shielding of the input bond pad by an underlying buried layer connected to the on-chip ground can result in severe coupling problems if the ground is noisy. In this case we must renounce this shield or connect it to another, less critical circuit node.

[18] As long as they are not part of a feedback path [as, e.g., in Fig. 2(a)].

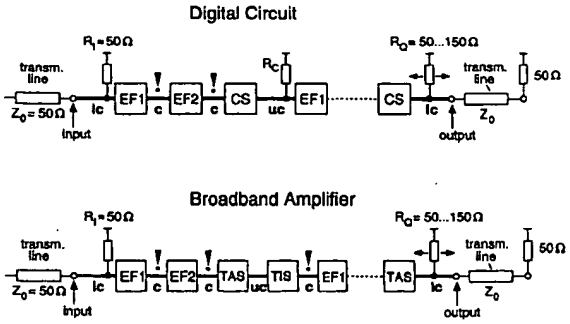


Fig. 10. Characterization of different signal-line segments in high-speed IC's, c = critical, lc = less critical, uc = uncritical. The input resistor R_I and especially the load resistor of the CS, R_C, are preferably located at the line end.

line inductances can result in severe ringing and even in undamped oscillations.[19] In addition to shortening the lines, it is sometimes helpful to use microstrip lines with comparatively wide stripes, possibly combined with a thinner dielectric layer (use of neighbored metal layers for the signal line and the ground plane), in order to reduce the effective line inductance (of course at cost of an increased line capacitance).

In Fig. 10 the input and output lines are characterized as "less critical," ranking between the critical and uncritical lines. Double arrows indicate that the on-chip output resistor R_Q can be alternatively located at both ends of the line, resulting in a somewhat different line influence on circuit behavior. In [10] and [11] a single EF is used at the amplifier input in order to reduce the equivalent input noise voltage density.

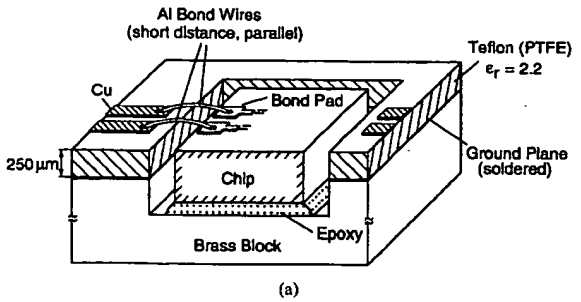
There are a lot of other examples where on-chip wiring can lead to severe performance degradation, especially in low-ohmic circuits. However, their discussion would be beyond the scope of this paper.

VI. Mounting Aspects

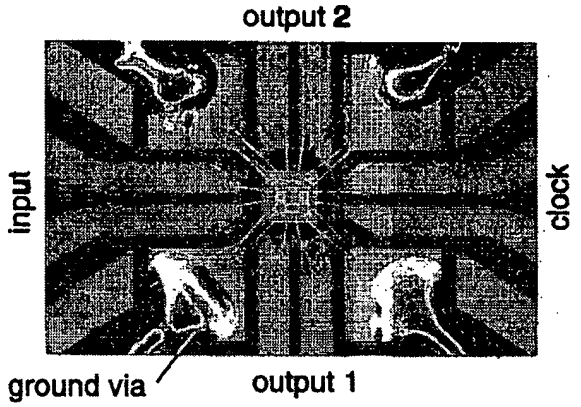
First, it has to be pointed out that on-wafer measurements, which are often presented in publications, are useful for preselection of the chips but not at all for the final characterization of a circuit. The reason for this is that in practice only mounted chips can be used and that, therefore, the mounting parasitics have to be considered already in the circuit design (cf. Fig. 1).

For measurements, our chips are directly glued into a measuring socket and are conventionally ultrasonic bonded using 1-mil aluminum wires. The socket consists of a brass block, a dielectric substrate plate with 50-Ω microstrip lines, and SMA or K connectors. The substrate, which is soldered on the brass block, is either a ceramic plate with gold clad or a Teflon plate (PTFE) with copper clad on both sides. If possible, we prefer the latter substrate because it is less expensive and can easily be treated [5], [18]. Vias to the ground plane can be made easily by use of copper ribbons

[19]For example, in the case of the laser/modulator driver of Table I, line inductances as low as 20 pH behind each of the three EF's driving the output current switch [cf. Fig. 12(a)] would result in undamped oscillations.



(a)



(b)

Fig. 11. Measuring socket: (a) schematic section and (b) photograph of a mounted DEMUX chip with ground vias in each corner (chip size: 0.74 mm × 0.74 mm).

soldered on both substrate sides. A section of the measuring socket and a practical example of a mounted chip (32 Gb/s DEMUX [6]) is shown in Fig. 11.

The parasitic bond inductances (as far as they are undesired) are reduced by the following measures (cf. Fig. 11).

- By milling a recess for the chip into the brass block the surfaces of the chip and the microstrip lines are at the same level, resulting in short bond wires.
- In each corner of the chip there are ground pads which are all connected by short bond wires to adjacent pads (vias) on the substrate. If the ground pads on the chip must be more concentrated for some reasons,[20] multiple bonding pads should be used, providing a sufficient angle between the directions of the bond wires.
- For differential operation, the effective (odd) bond inductance at the (complimentary) signal pads can be essentially reduced by use of parallel bond wires with small distance between them. This is because the mutual inductance increases. Formulas for the calculation of this

[20]For example, in the case of single-ended amplifiers, short bond wires are required not only to ground but also to external decoupling capacitors. These capacitors have to be arranged close to the chip so that ground bonding to all four corners of the chip [as in Fig. 11(b)]may no longer be possible.

effect are, e.g., given in [41]. They proved to be well suited for practical applications [18].

As a problem of the low-permittivity substrate, 50-Ω lines have comparatively wide stripes (750 μm). Therefore, near the chip, narrower stripes must be used. However, for complimentary lines the (differential) characteristic impedance of 50 Ω can be maintained by reducing the distance between both lines [cf. Fig. 11(b)].

The suitability of this mounting technique up to very high data rates was confirmed by numerous examples. For example, the 50 Gb/s time-division multiplexer, presented in the next section, was mounted in the same way as shown in Fig. 11 with a single exception: the semirigid cables at the (50 Gb/s) output were directly soldered with the microstrip lines of the substrate, thus eliminating the connectors between them. This simple connection shows an excellent high-frequency performance.

VII. Some Circuit Examples and Measuring Results

Some high-speed circuits designed by applying the techniques discussed in this paper have already been presented in Table I of the Introduction. In this section, three rather different kinds of circuits are selected, a 14 Gb/s driver circuit, a 13 Gb/s AGC broadband amplifier, and a 50 Gb/s MUX. For a fair comparison of the speed performance of these IC's, it has to be considered that the first two were fabricated in (different) production technologies, while for the third one, an advanced laboratory technology was used.

A. Driver Circuit

Fig. 12(a) shows the (simplified) diagram of a driver circuit developed for the transmitter of a 10 Gb/s optical-fiber link [8]. It was fabricated with the Siemens production technology B6HF [13]. Its single-ended output is able to drive both a laser diode directly or an external modulator. While in the first case the output current swing has to be varied within a wide range (here: 15 mA $\leq \Delta I_Q \leq$ 60 mA) a comparatively high voltage swing (here: $\Delta V_Q = 3.6\ V_{p-p}$) is required in the second case. The circuit consist of two current switches, which are (for the reasons discussed in Section III-A) each driven by three EF pairs. The basic problems of such a circuit and their solutions discussed in [8] shall briefly be mentioned.

—*Asymmetrical pulse shape due to the single-ended output.* As a consequence, the eye diagrams can be severely degraded. By introduction of two offset voltages [caused by the currents I_{OF1} and I_{OF2} in Fig. 12(a)] symmetrical output pulses can be generated.

—*Large output time constant due to the high voltage swing* ($\tau_Q \propto \Delta V_Q$). Here, the only practical way to achieve the required quality of the output eye diagrams at 10 Gb/s and above was to steepen the pulse edges by a peaking bond inductor L_P (≈2 nH) at the output (cf. Section III-C).

—*Wide output current range for direct laser modulation.* For this, in addition to the current source of the output stage, the currents through several other stages have to be optimally varied via a single control input CON in order to get clear eye diagrams.

—*Exceeding the breakdown voltage* V_{CE0}(≈3.7 V). This is a consequence of the high output voltage swing. The resulting maximum V_{CE} of 5.5 V can, however, be tolerated in the present case (cf. Section IV-C).

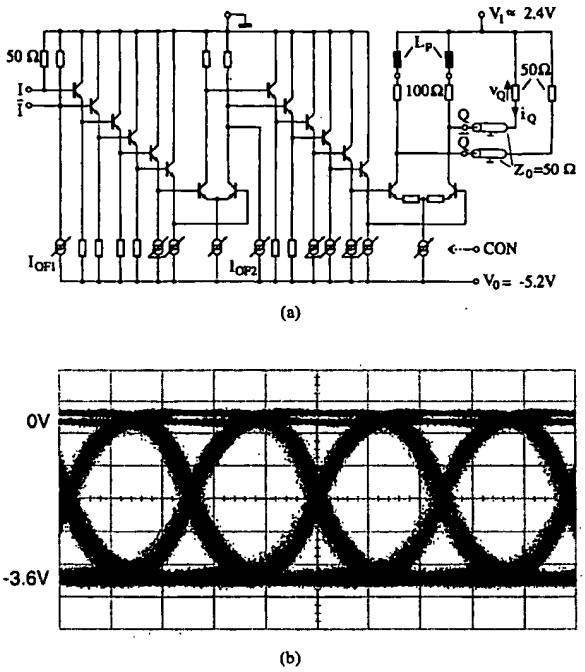
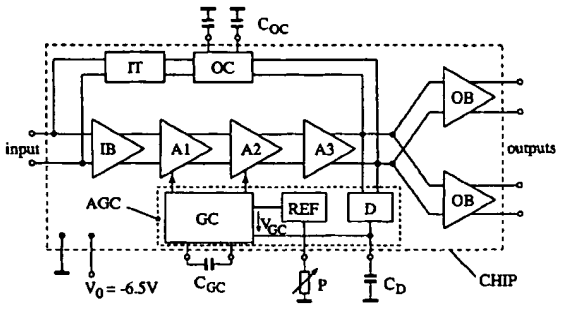


Fig. 12. Laser/modulator driver. (a) Circuit diagram. (b) Output eye diagram at 14 Gb/s (29 ps/div.).

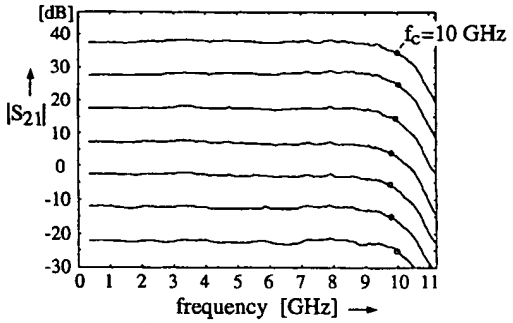
The measured (single-ended) output eye diagram at 14 Gb/s is shown in Fig. 12(b). In this case, the current switched by the output transistors is as high as 108 mA, resulting in an output voltage swing of $\Delta V_Q = 3.6\ V_{p-p}$ (across an external load of $R_L = 50\ \Omega$).

B. AGC Broadband Amplifier

Based on the amplifier cells discussed in Section III-B (Fig. 5) a main amplifier with automatic gain control (AGC) was developed for a 10 Gb/s optical-fiber link [10] and fabricated with an advanced version of Motorola's production technology MOSAIC V [42]. The block diagram of the dc coupled and fully differential circuit is given in Fig. 13(a). The h.f. part consists of an high-ohmic input buffer IB, two controllable amplifier cells A1 (cf. Fig. 5) and A2 (modified compared to A1 [10]), a constant-gain cell A3, and two separated 50-Ω output buffers. The driving transmission line is well matched by a 50-Ω input termination circuit IT which is combined with an offset-control circuit OC. For gain control, the output signal of A3 is detected by a peak detector D and compared with the nominal voltage (REF). The voltage

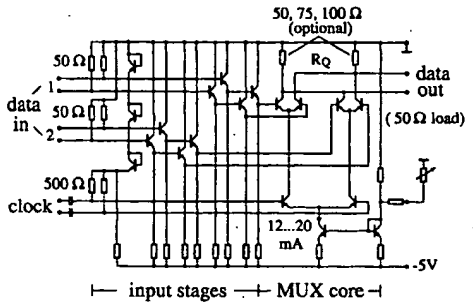


(a)

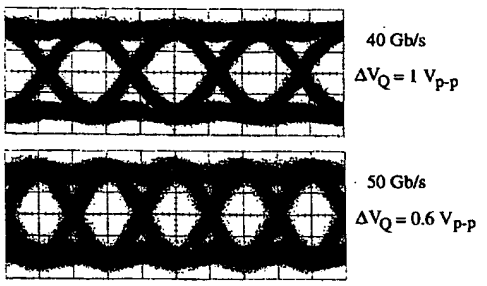


(b)

Fig. 13. Broadband AGC amplifier with two outputs. (a) Block diagram. (b) Gain versus frequency characteristics for a wide range of gain.



(a)



(b)

Fig. 14. 2:1 MUX. (a) Circuit diagram. (b) Output eye diagrams at 40 and 50 Gb/s (10 ps/div.).

difference V_{GC} is amplified by a high-gain dc amplifier which generates the two different control voltages for A1 and A2.

Aspects for the design of this amplifier have already been discussed in [10]. Besides the high data rate ($\geq$10 Gb/s) and the high gain (S_{21} = 37 dB) also the demands on high input dynamic range (41 dB) and sufficiently low equivalent noise voltage density at the input (2.5 nV/$\sqrt{\text{Hz}}$) had to be met. Fig. 13(b) shows the measured frequency response of the gain ($|S_{21}|$) with the low-frequency gain as a parameter.[21] Despite the extremely wide gain range (60 dB!) the 3-dB cutoff frequency f_C is nearly constant ($\approx$10 GHz). Moreover, clear eye diagrams were measured up to 13 Gb/s at a constant output voltage swing of 500 mV_{p-p} within the total input dynamic range [10].

C. Time-Division Multiplexer

The circuit diagram of the 2:1 MUX, well proven in previous designs [5], is shown in Fig. 14(a) [4]. No separate output buffer is used as it would decrease the operating speed and increase the power consumption. In another version, a clock input stage with 50-Ω on-chip matching resistors is used [cf. Fig. 2(b)], [6].

[21] In this case the automatic-gain-control had been switched-off and V_{GC} was varied via an external control voltage rather than by the output of the peak detector.

After fabrication in a laboratory technology [14], the chip was mounted on the measuring socket presented in Section VI. The 25 Gb/s input signals for driving the circuit were generated by multiplexing four pseudorandom pulse sequences with a word length of $2^{15} - 1$ b each [43]. The output eye diagrams measured at 40 and 50 Gb/s are shown in Fig. 14(b) with (differential) voltage swings of $\Delta V_Q = 1\ V_{p-p}$ and 0.6 V_{p-p}, respectively [4]. To the best of the authors' knowledge, 50 Gb/s is the highest data rate ever achieved with an integrated circuit in any technology.

VIII. Conclusion

Design aspects for digital and analog Si-bipolar IC's operating well above 10 Gb/s were presented and their usability was confirmed by measurements of practical IC's. Data rates up to 50 and 46 Gb/s were achieved with a MUX and DEMUX, respectively. However, we cannot conclude from these results that present Si technologies and design methods are already sufficient for all the other circuits (especially for the amplifiers and laser/modulator driver) in 40 Gb/s optical communication systems just under development [22]. To reach this goal, a further improvement not only of the technology but also of the mounting technique and of the models for transistor and circuit parasitics are required. In addition, the specification of the most critical circuits in 40 Gb/s systems must be relaxed compared to those given in Table I for 10 Gb/s systems. However, from their present designs, the authors expect that

the circuits principles and design aspects presented in this paper remain suitable also for such a high operating speed.

ACKNOWLEDGMENT

The authors would like to thank all members of their group for the excellent design work as well as their industrial partners, especially the staff members of Siemens and ANT Nachrichtentechnik, for the successful cooperation.

REFERENCES

[1] H.-M. Rein, "Design aspects of 10 to 40 Gb/s digital analog Si-bipolar IC's," in *Proc. Symp. on VLSI Circuits (invited paper)*, Kyoto, June 1995, pp. 49–54.

[2] ——, "Very-high-speed Si and SiGe bipolar IC's," in *Proc. ESSDERC'95*, The Hague, The Netherlands, Sept. 1995, pp. 45–56.

[3] A. Felder, M. Möller, J. Popp, J. Böck, and H.-M. Rein, "46 Gb/s DEMUX, 50 Gb/s MUX, and 30 GHz static frequency divider inn silicon bipolar technology," *IEEE J. Solid-State Circuits*, vol. 31, no. 4, Apr. 1996, pp. 481–486.

[4] M. Möller, H.-M. Rein, A. Felder, J. Popp, and J. Böck, "50 Gb/s time-division multiplexer in Si-bipolar technology," *Electron. Lett.*, vol. 31, pp. 1431–1433, Aug. 1995.

[5] H.-M. Rein, J. Hauenschild, M. Möller, W. McFarland, D. Pettengill, and J. Doernberg, "30 Gbit/s Si multiplexer and demultiplexer IC's in silicon bipolar technologies," *Electron. Lett.*, vol. 28, no. 1, pp. 97–99, Jan. 1992.

[6] J. Hauenschild, A. Felder, M. Kerber, H.-M. Rein, and L. Schmidt, "A 22 Gb/s decision circuit and a 32 Gb/s regenerating demultiplexer IC fabricated in silicon bipolar technology," in *Proc. IEEE 1992 BCTM*, Minneapolis, Oct. 1992, pp. 151–154.

[7] A. Felder, J. Hauenschild, M. Kerber, and H.-M. Rein, "Static silicon frequency divider for low power consumption (4 mW, 10 GHz) and high speed (160 mW, 19 GHz)," in *Proc. IEEE 1992 BCTM*, Minneapolis, Oct. 1992, pp. 159–162.

[8] H.-M. Rein, R. Schmid, P. Weger, T. Smith, T. Herzog, and R. Lachner, "A versatile Si-bipolar driver circuit with high output voltage swing for external and direct modulation of laser diodes in 10 Gb/s optical fiber links," *IEEE J. Solid-State Circuits*, vol. 29, pp. 1014–1021, Sept. 1994.

[9] M. Neuhäuser, H.-M. Rein, H. Wernz, and A. Felder, "A 13 Gbit/s Si bipolar preamplifier for optical front ends," *Electron. Lett.*, vol. 29, no. 5, pp. 492–493, Mar. 1993.

[10] M. Möller, H.-M. Rein, and H. Wernz, "13 Gb/s Si-bipolar AGC amplifier with high gain and wide dynamic range for optical-fiber receivers," *IEEE J. Solid-State Circuits*, vol. 29, no. 7, pp. 815–822, July 1994.

[11] ——, "15 Gb/s high-gain limiting amplifier fabricated in a Si-bipolar production technology," *Electron. Lett.*, vol. 30, no. 18, pp. 1519–1520, Sept. 1994.

[12] A. Felder, R. Stengl, J. Hauenschild, H.-M. Rein, and T. F. Meister, "25 to 40 Gb/s IC's in a 0.8 μm silicon selective epitaxy bipolar technology," in *Dig. 1993 IEEE ISSCC*, San Francisco, pp. 156–157, 281, Feb. 1993.

[13] H. Klose *et al.*, "B6HF: A 0.8 micron 25 GHz/25 ps bipolar technology for 'mobile radio' and 'ultra fast data link' IC-products," in *Proc. IEEE 1993 BCTM*, Minneapolis, Oct. 1993, pp. 125–127.

[14] J. Böck et al., "A 0.6 μm Si bipolar technology with 17 ps CML gate delay and 30 GHz static frequency divider," in *Proc. ESSDERC'95*, The Hague, The Netherlands, Sept. 1995, pp. 418–420.

[15] T. F. Meister *et al.*, "Sub-20 ps silicon bipolar technology using selective epitaxial growth," in *Dig. IEEE IEDM 1992*, San Francisco, pp. 401–404, Dec. 1992.

[16] A. Schüppen, A. Gruhle, U. Erben, H. Kibbel, and U. König, "Multi emitter finger SiGe-HBT's with f_{max} up to 120 GHz," in *Dig. IEEE IEDM 1994*, San Francisco, Dec. 1994, pp. 377–380.

[17] J. Hauenschild and H.-M. Rein, "Influence of transmission-line interconnections between Gbit/s IC's on time jitter and instabilities," *IEEE J. Solid-State Circuits*, vol. 25, no. 3, pp. 763–766, June 1990.

[18] J. Hauenschild, "Optimization of signal transmission between integrated silicon-bipolar circuits in the Gbit/s range," Dr.-Ing. thesis (in German), Ruhr-University Bochum, Fakultät für Elektrotechnik, Bochum, Germany, 1993.

[19] E. M. Cherry and D. E. Hooper, "The design of wide-band transistor feedback amplifier," in *Inst. Elec. Eng. Proc.*, Feb. 1963, vol. 110, no. 2, pp. 375–389.

[20] M. Neuhäuser, M. Möller, H.-M. Rein, and H. Wernz, "Low-noise, high transimpedance Si-bipolar AGC amplifier for 10 Gb/s optical-fiber links," *IEEE Photonics Technol. Lett.*, vol. 7, no. 5, pp. 549–551, May 1995.

[21] M. Neuhäuser, H.-M. Rein, and H. Wernz, "Low-noise, high-gain Si-bipolar preamplifiers for 10 Gb/s optical-fiber links—Design and realization," *IEEE J. Solid-State Circuits*, vol. 31, pp. 24–29, Jan. 1996.

[22] "Photonik II. Joint research program supported by the German Ministry of Research and Technology (BMBF)".

[23] M. Ishihara, O. Nakajima, H. Ichino, and Y. Yanauchi, "9 GHz bandwidth, 8–20 dB controllable-gain monolithic amplifier using AlGaAs/GaAs HBT technology," *Electron. Lett.*, pp. 1317–1318, Sept. 1989.

[24] J. L. Hullett, H. B. Doan, and G. Rosman, "A modified receiver for optical transmission systems," *IEEE Trans. Commun.*, pp. 1514–1518, Dec. 1975.

[25] R. Ranfft and H.-M. Rein, "High-speed bipolar logic circuits with low power consumption for LSI—A comparison," *IEEE J. Solid-State Circuits*, vol. 17, no. 4, pp. 703–712, Aug. 1982.

[26] C. T. Kirk, "A theory of transistor cutoff frequency (f_T) falloff at high current densities," *IEEE Trans. Electron Devices*, vol. 9, pp. 164–174, Mar. 1962.

[27] H. Stübing, H.-M. Rein (part I) and respectively, H.-M. Rein, M. Schröter (part II), "A compact physical large-signal model for high-speed bipolar transistors at high current densities, part I: One-dimensional model, part II: Two-dimensional model and experimental results," *IEEE Trans. Electron Devices*, vol. 34, no. 8, pp. 1741–1761, Aug. 1989.

[28] M. Schröter and H.-M. Rein, "Transit time of high-speed bipolar transistors in dependence on operating point, technological parameters and temperature," in *Proc. IEEE 1989 BCTM*, Minneapolis, Sept. 1989, pp. 250–253.

[29] D. J. Walkey, M. Schröter, and S. Voinigescu, "Predictive modeling of lateral scaling in bipolar transistors," in *Proc. IEEE 1995 BCTM*, Minneapolis, Oct. 1995, pp. 74–77.

[30] N. Itoh, Y. Yoshida, S. Watanabe, Y. Katsumata, and H. Iwai, "The analysis of silicon bipolar transistor scaling-down scheme for low noise and low power analog application," in *Proc. IEEE 1994 BCTM*, Minneapolis, Oct. 1994, pp. 60–63.

[31] M. Schröter and H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar IC's using both a new compact model and a device simulator," *IEEE J. Solid-State Circuits*, vol. 30, no. 5, pp. 551–562, May 1995.

[32] M. Schröter, "A survey of present compact models for high-speed bipolar transistors," *Frequenz*, vol. 47, pp. 178–190, July/Aug. 1993.

[33] H.-M. Rein and M. Schröter, "TRADICA—A computer program for dimensioning and calculating high-speed integrated transistors," Ruhr-University Bochum, Germany, 4.1 Edition, *User's Manual*, 1993.

[34] M. Schröter, "Physical models for high-speed silicon bipolar transistors—A comparing survey," Ruhr-University Bochum, Fakultät für Elektrotechnik, Bochum, Germany, Habilitation (in German), 1993.

[35] A. Koldehoff, M. Schröter, and H.-M. Rein, "A compact bipolar transistor model for very-high-frequency applications with special regard to narrow emitter stripes and high current densities," *Solid-State Electron.*, vol. 36, no. 7, pp. 1035–1048, July 1993.

[36] M. Pfost, H.-M. Rein, and T. Holzwarth, "Modeling of the substrate effect in high-speed Si-bipolar IC's," in *Proc. IEEE 1995 BCTM*, Minneapolis, Oct. 1995, pp. 182–185.

[37] H.-M. Rein, T. Schad, and R. Zühlke, "The influence of the base resistance and the carrier avalanche effect on the output characteristics of planar transistors," *Solid-State Electron.*, vol. 15, pp. 481–500, 1972.

[38] H.-M. Rein, "Calculation of the breakdown characteristics of bipolar transistors using a two-dimensioned model," *Solid-State Electron.*, vol. 19, pp. 145–147, 1976.

[39] E. Zanoni *et al.*, "A physics-based, accurate SPICE model of impact-ionization effects in bipolar transistors," in *Proc. ESSDERC'94*, Edinburgh, Sept. 1994, pp. 181–184.

[40] R. Reimann and H.-M. Rein, "A single-chip bipolar AGC amplifier with large dynamic range for optical-fiber receivers operating up to 3 Gbit/s," *IEEE J. Solid-State Circuits*, vol. 24, no. 6, pp. 1744–1748, Dec. 1989.

[41] H. Meinke and F. W. Gundlach, *Taschenbuch der Hochfrequenztechnik*. Berlin: Springer, 1986.

[42] V. dela Torre *et al.*, "MOSAIC V—A very high performance bipolar technology," in *Proc. IEEE 1991 BCTM*, Minneapolis, Sept. 1991, pp. 21–24.

[43] M. Bussmann, U. Langmann, W. J. Hillary, and W. W. Brown, "PRBS generation and error detection above 10 Gb/s using a monolithic Si bipolar IC," *IEEE J. Lightwave Technol.*, vol. 12, no. 2, pp. 353–359, Feb. 1994.
[44] J. Te Winkel, "Extended charge-control model for bipolar transistors," *IEEE Trans. Electron Devices*, vol. ED-20, pp. 389–394, Apr. 1973.
[45] M. Pfost, H.-M. Rein, and T. Holzwarth, "Modeling substrate effects in the design of high-speed Si-bipolar IC's," to be published, *IEEE J. Solid-State Circuits*, vol. 31, 1996.

H.-M. Rein (M'92) received the Dipl.-Ing. and Dr.Ing. degrees in electrical engineering from the University of Stuttgart, Germany, in 1964 and 1968, respectively.

From 1968 to 1973, he worked in the field of integrated circuit techniques for the Semiconductor Division of AEG-Telefunken in Heilbronn, Germany. He was head of the departments "Basic Development of IC's" and "MOS Circuits." In 1973, he became a Professor at Ruhr-University, Bochum, Germany, where he is now teaching integrated circuits. Presently, his main fields of research are analog and digital high-speed bipolar IC's, e.g., for optical communication systems, and their fundamental limits, optimization techniques for integrated circuits, and modeling of very fast bipolar transistors. He is author and coauthor of more than 100 technical publications and of a textbook on integrated bipolar circuits.

M. Möller, for a photograph and biography, see p. 486 of the April issue of this JOURNAL.